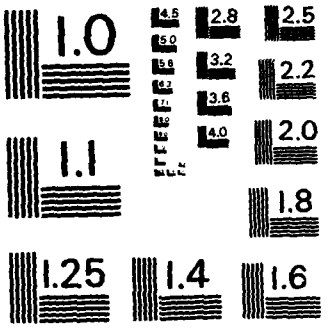
AD-A159 784 A SIMPLE UNCONSTRAINED DUAL CONVEX PROGRAMMING METHOD FOR THE COMPUTATION..(U) TEXAS UNIV AT AUSTIN CENTER FOR CYBERNETIC STUDIES P BROCKETT ET AL. MAY 84 1/1

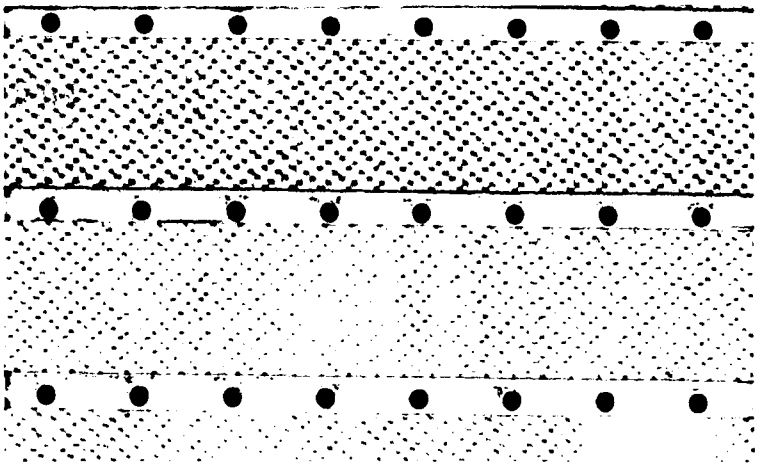
UNCLASSIFIED CCS-492 N00014-82-K-0295 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS - 1963 - A

AD-A159 784

(2)



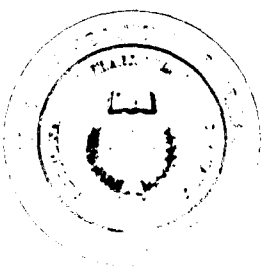
CENTER FOR CYBERNETIC STUDIES

The University of Texas
Austin, Texas 78712

DTIC
ELECTE
OCT 1 1985
D
B

DTIC FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited



85 10 01 112

Research Report CCS 492

A SIMPLE UNCONSTRAINED DUAL CONVEX PROGRAMMING METHOD FOR THE COMPUTATION OF DISCRETE MAXIMUM ENTROPY DISTRIBUTIONS

by

P. Brockett
K. Paick

May 1984

DTIC
ELECTE
OCT 1 1985
S D
B

This research was supported in part by ONR Contracts N00014-82-K-0295 and N00014-81-C-0236 with the Center for Cybernetic Studies, T... University of Texas at Austin. Reproduction in whole or in part is permitted for any purpose of the United States Government.

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
College of Business Administration, CBA 5.202
The University of Texas at Austin
Austin, Texas 78712
(512) 471-1821

(-A-)

A SIMPLE UNCONSTRAINED DUAL CONVEX PROGRAMMING METHOD FOR THE COMPUTATION OF DISCRETE MAXIMUM ENTROPY DISTRIBUTIONS

by

P. Brockett and K. Paick

ABSTRACT

This document
~~We~~ formulates the generalized constrained maximum entropy problem often used in a decision making context as an extended dual convex programming problem. We then present the dual problem, is then presented. In this dual setting the primal Lagrange multipliers are precisely the dual variables, and are easily calculated directly by virtue of the simple structure of the dual problem. An example involving the selection of best equipment for an oil spill is presented as an illustration. ~~We~~ The authors contrast ~~our~~ their solution with those given by previous authors.

KEY WORDS

Maximum entropy computation

Dual program for maximum entropy problems

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist Special
A-1

DTIC COPY INSPECTED

The notion of information is strongly connected to the amount of uncertainty. In many problems encountered in operations research practice, it is useful to estimate the discrete probability distribution for a random phenonmenon under uncertainty. The general maximum entropy principle is a very useful method for incorporation the uncertainty of some situation into a probability distribution where one is trying to make the most out of some limited knowledge and resources.

The maximum entropy estimation, a special case of minimum discrimination information, has been used in numerous fields (e.g. Brockett et al [1984], Thomas [1979].

In a recent paper, Freund and Saxena [1984] gave an algorithm to compute maximum entropy probability estimates. In this paper we present a much more general, and much easier computational method for obtaining these estimates. Additionally, our method easily extends to the computation of minimum discrimination information estimates as well. The bulk of this paper centers on the development of a dual convex programming formulation for maximum entropy estimation, and shows how to view maximum entropy estimation from this dual convex programming point of view. We then point out the analytical properties of the estimates which follow directly from the form of this duality. Section I contains the mathematical formulation of maximum entropy. In section II we present the unconstrained dual formulation for Lagrange multipliers due to Charnes and Cooper [1975] and Charnes, Cooper and Seiford [1978]. Section III contains an application of the unconstrained dual convex

programming method to an oil spill equipment selection problem considered in Thomas [1979], and also in Freund and Saxena [1984]. The algorithm we present is more general and easier computationally than that given in Freund and Saxena [1984].

I. MAXIMUM ENTROPY

Mathematically the problem of maximum entropy estimation is to determine that density function p which is maximally uncertain, and which satisfies certain given constraints, e.g.

$$\max H(p) = -\int p(x)\ln[p(x)]\lambda(dx)$$

$$\text{s.t.} \quad \int h_0(x)p(x)\lambda(dx) = \theta_0 = 1$$

$$\int h_1(x)p(x)\lambda(dx) = \theta_1$$

$$\vdots$$

$$\int h_k(x)p(x)\lambda(dx) = \theta_k$$

Here λ is some dominating measure for p (usually Lebesgue measure in the continuous case, or counting measure in the discrete case), $\theta_1,\ldots,\theta_k$ are the given constant values for a known set of moment functions $h_1,\ldots,h_k$ and $h_0(x)=1$. Frequently in hypothesis testing or estimation of an inferential distribution by maximum entropy estimation, one has information about the possible candidate distribution in the form of inequality constraints in addition to equality constraints. The form of range constraints of probability distribution can be easily transformed to inequality constraints. In this case we can add following constraints to the original constraints;

$$\int h_{k+1}(x)p(x)\lambda(dx) \leq \theta_{k+1}$$
$$\vdots$$
$$\int h_{k+n}(x)p(x)\lambda(dx) \leq \theta_{k+n}.$$

The explicit calculation of maximum entropy density subject to the given constraints is carried out by Lagrange multipliers. Since solving the maximum entropy estimate entails solving the highly non-linear constraint equation, it has been difficult to solve the Lagrange multiplier system explicitly in order to obtain a closed form solution expressed directly in terms of the known expected values θ_i (Leblanc and Reisher [1981], Brockett et al [1980]). For this reason certain numerical solutions were derived by approximation (first order approximation by Guiasu [1980], second order approximation by Leblanc and Riesher [1981]). Furthermore, the solutions in Thomas [1980] and Freund and Saxena [1984] turn out to be different from the optimal solutions. We will discuss these examples in section III.

II. UNCONSTRAINED DUAL PROGRAMMING APPROACH TO ESTIMATION.

In the first part of this section, we shall present maximum entropy estimation in the discrete case via dual convex programming with only non-positivity constraints. These results are special cases of the results given in Charnes and Cooper[1975] and Charnes, Cooper, and Seiford[1978].

Theorem 2.1

The following linear constrained maximum entropy primal problem

$$\sup \nu(\delta) = -\delta^t \ln\delta$$

$$\text{s.t. } \delta^t A^1 = b^{1t} \qquad (P)$$

$$\delta^t A^2 \leq b^{2t}$$

$$\delta \geq 0$$

has a dual problem

$$\inf \xi(z) = \exp(A^1 z^1 + A^2 z^2) - b^{1t} z^1 - b^{2t} z^2$$

$$\text{s.t.} \qquad z^2 \leq 0. \qquad (D)$$

There are three mutually exclusive and collectively exhaustive duality states;

(1) $\Delta \equiv \{\delta : \delta^t A^1 = b^{1t},\ \delta^t A^2 \leq b^{2t},\ \delta \geq 0\} = \phi$ and $\xi(z)$ is unbounded below.

(2) Every feasible solution of (P) has a zero component for all $\delta \in \Delta \neq \phi$ and $\xi(z)$ with non-positive z^2 has only an infimum. In this case $\inf \xi(z) = \max \nu(\delta) = \min \xi_D(z)$ where $\xi_D(z)$ contains only those terms of $\xi(z)$ for which $\delta_i > 0$ in some $\delta \in \Delta$.

(3) There exists $\delta \in \Delta$ with $\delta > 0$ and $\xi(z)$ has a minimum at z^*. In this case following relationships obtain between the optimal primal and dual variables

a) $\inf \xi(z^*) = \sup \nu(\delta^*) = \max \nu(\delta^*) = \min \xi(z^*)$

b) $\nu(\delta)$ has a unique maximum at $\delta^* > 0$

c) $\delta^{*t} = \exp[A^1 z^{1*} + A^2 z^{2*}]$.

Note that state (3) is the usual state considered in applied problems.

Proof (adapted from Charnes, Cooper and Seiford [1978])

The constraints in primal problem may be written as

$$\delta^t A^1 = b^{1t}$$
$$\delta^t A^2 + \gamma^t = b^{2t} \qquad (2.1)$$
$$\delta,\ \gamma \geq 0$$

Here A^1 is $m_1 x n_1$ and A^2 is $m_2 x n_2$. By the duality inequality,

$$-\sum_i \delta_i \ln\delta_i \leq \sum_{i\in A} (\exp(x_i) - \delta_i x_i) - \sum_{i\in B} \gamma_i y_i$$

with (2.1), $y_i \leq 0$, $i\in B$. Also $A=\{1,\ldots,m_1\}$, and $B=\{m_1+1,\ldots,m_2\}$. i.e.

$-\delta^t \ln\delta \leq \min \exp(x) - (\delta^t x + \gamma^t y) \equiv K(\delta,\gamma, x,y)$ with $\delta,\gamma \geq 0$, $y\leq 0$.

To decouple, we obtain $\delta^t x + \gamma^t y = b^{1t} z^1 + b^{2t} z^2$ and

$$(\delta^t,\gamma^t)\begin{bmatrix} x \\ y \end{bmatrix} = (\delta^t,\gamma^t)\begin{bmatrix} A^1 & A^2 \\ 0 & I \end{bmatrix}\begin{bmatrix} z^1 \\ z^2 \end{bmatrix}$$

where we have set

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A^1 & A^2 \\ 0 & I \end{bmatrix}\begin{bmatrix} z^1 \\ z^2 \end{bmatrix}$$

thus we obtain the dual problem

$$\inf \xi(z) \equiv \exp(A^1 z^1 + A^2 z^2) - b^{1t} z^1 - b^{2t} z^2$$
$$\text{s.t.} \quad z^2 \leq 0 .$$

Because state (3) is the most usual and encountered state, we shall present the proof for state (3) only. The proof of (1) and (2) may be found in Charnes, Cooper and Seiford [1978].

Let $K(\delta,\gamma,x,y) = \exp(x) - \delta^t x - \gamma^t y$ for $\delta,\gamma\geq 0$, $x\in R$ and define $g(\delta) = \inf K(\delta,\gamma, x,y) = -\delta^t \ln\delta$. Because of the constraint

$$(\delta^t,\gamma^t)\begin{bmatrix} A^1 & A^2 \\ 0 & I \end{bmatrix} = (b^{1t}, b^{2t}) \text{ and by setting } \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A^1 & A^2 \\ 0 & I \end{bmatrix}\begin{bmatrix} z^1 \\ z^2 \end{bmatrix}$$

we have $g(\delta) = -\delta^t \ln(\delta) \leq \exp(x) - \delta^t x - \gamma^t y$

$= \exp[A^1 z^1 + A^2 z^2] - \delta^t [A^1 z^1 + A^2 z^2] - \gamma^t z^2$

$= \exp[A^1 z^1 + A^2 z^2] - b^1 z^1 - b^2 z^2$

which holds for all $z^1, z^2 \leq 0$, $\delta = \{\delta: \delta^t A^1 = b^{1t}, \delta^t A^2 \leq b^{2t}, \delta \geq 0\}$, and $\gamma = \{\gamma : \delta^t A^2 + \gamma^t = b^{2t}, \gamma \geq 0\}$. The equality holds for $\delta^* = \exp [A^1 z^{1*} + A^2 z^{2*}]$. Q.E.D.

If the requisites for state determination are not obvious, the state may be charaterized by means of the following linear programming problem:

$$\begin{aligned} &\max \ \mu \\ &\text{s.t. } \mu w^t - \delta^t \leq 0 \\ &\qquad \delta^t A^1 = b^{1t} \\ &\qquad \delta^t A^2 \leq b^{2t} \\ &\qquad \delta \geq 0 \end{aligned}$$

where $w^t = (1,1,\ldots,1)$. State (1) corresponds to infeasibility, state (2) corresponds to $\mu^* = 0$ and state (3) corresponds to $\mu^* > 0$. It is obvious that there is no linear independence requirement, and all possible behaviors for the system Δ are considered.

This result is very attractive since the dual problem (D) is an convex programming problem involving only exponential and linear terms with non-positivity constraint for z^2. Moreover, the desired Lagrangian multipliers for the maximum entropy estimate (P) are precisely the dual variables to (D), and (D) is easily solved numerically because of the simply constrained nature of the problem (even unconstrained in the

equality constrained case of the primal). Any of a number of readily available non-linear programming codes can be used to solve the dual formulation. Moreover, since we explicitly know the parametric form of the optimizing density in terms of the unknown Lagrange multipliers, and this form is unique and continuous in the unknown paramenters, the procedure we employ in obtaining the Lagrange parameters via the dual convex programming problem and then substituting into the parametric form is stable numerically.

An alternative structure for a dual problem is a single one parameter sequence of equality form. In order to make $K(\delta,\gamma, x,y)$ more symmetric in δ and γ, and to remove the restriction of $y\leq 0$, we adopt the same procedure we employed before. We can change (P) into (P') w.l.o.g..

$$(P') \quad \max \ \delta^t \ln\delta - \varepsilon^t \ln\gamma$$

$$\text{s.t. } \delta^t A^1 = b^{1t}$$

$$\delta^t A^2 + \gamma^t = b^{2t}$$

$$\delta, \gamma \geq 0 \text{ where } \varepsilon > 0.$$

Let define $K(\delta,\gamma,x,y) = \exp(x) + \varepsilon\exp(y/\varepsilon) - \delta^t x - \gamma^t y$ with $\varepsilon>0$. By the Charnes-Cooper duality theorem [1975], the following inequality holds.

$$-\delta^t \ln\delta - \gamma^t \ln\gamma \leq K(\delta,\gamma,x,y).$$

Because of the given constraint

$$(\delta^t, \gamma^t) \begin{bmatrix} A^1 & A^2 \\ 0 & I \end{bmatrix} = (b^{1t}, b^{2t}), \text{ and by setting}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A^1 & A^2 \\ 0 & I \end{bmatrix} \begin{bmatrix} z^1 \\ z^2 \end{bmatrix}$$

we have

$$-\delta^t \ln\delta \leq \exp(x) - \delta^t x - \gamma^t y + \varepsilon\exp(y/\varepsilon).$$

Thereby a new unconstrained form of dual problem (D') for (P') is obtained.

$$(D') \quad \inf \xi(z) \equiv \exp(A^1 z^1 + A^2 z^2) + \varepsilon\exp(z^2/\varepsilon) - b^1 t z^1 - b^2 t z^2$$

The duality theory of (P') and (D') is exactly that of the equality case presented in Brockett, Charnes and Cooper [1980]. Charnes, Cooper and Tyssedal [1983] proved that (P') is equivalent to (P) when ε approaches zero, and as a result (D') gives the solution to (D).

III. NUMERICAL EXAMPLES.

Two examples of maximum entropy estimates are presented. These are based on the oil spill problem in Thomas [1979]. The following four alternatives exemplify the decision problem for a particular habor area (see Thomas [1979] for the details)

a_1: contract all clean-up activities

a_2: procure equipment A for open area spills and contract for pierside clean-up

a_3: contract for open area spills and procure equipment set B for pierside clean-up

a_4: procure equipment set C for all spills.

Once the maximum entropy distribution is derived, the expected annual cost of each alternative j, $E(A_j)$, for both problems can be easily-calculated as $E(A_j) = \sum_i AC_{ij} p_i$ where AC_{ij} is the annual cost of j^{th} alternative for i^{th} state. We will provide the problem and solution for maximum entropy estimation part only.

The problem presented in Thomas [1979] reduces to the following:

$$\max \; -\Sigma p_i \ln p_i$$

s.t. $\Sigma p_i = 1$

$p_i \leq p_1 \quad i=3,\ldots,12$

$p_1+p_2 \geq p_3+p_5+p_6$

$p_5 \leq p_6$

$p_{10} \leq p_7$

$p_9 \leq p_8$

$p_{12} \leq p_{11}+p_6$

$p_7 \geq p_{10}+p_{11}+p_{12}$

$p_i \geq 0$

$.25 \leq p_1 \leq .60$

$.25 \leq p_2 \leq .50$

$.10 \leq p_3 \leq .40$

$0 \leq p_i \leq .30 \quad i=4,11,12$

$0 \leq p_i \leq .50 \quad i=5,\ldots,9$

$0 \leq p_{10} \leq .20$

Freund and Saxena simplified the above problem by choosing only a subset of the given constraints, namely the interval constraints, and non-negativity constraints for the p_i's and of course the normalizing constraint. Instead solving the simplified problem of Freund and Saxena by the technical algorithm they present, we note that we can get the optimal solution by intuition in this case. Due to the maximum entropy principle, p would be a uniform distribution if there were no constraints other than the usual normalizing and non-negativity constraints. p_1, p_2, and p_3 must, however, have the value of their respective lower bounds since these lower bound values are greater than 1/12 which is the value in uniform distribution giving maximum entropy. Given these lower bounds, the rest of the probabilities p_i, $i=4,\ldots,12$ would strive to uniformaly allocate the residual probability $1-p_1-p_2-p_3$.

Thus the intuitive optimal solution should be $p_1=p_2=.25$, $p_3=.10$ and $p_i=.4/9(i=4,\ldots,12)$. In fact, exactly the same optimal solution is obtained by solving the dual problem according to the technique given in this paper. The maximum entropy value is 2.1688. Freund and Saxena did not provide the probability distribution. The expected annual cost, however, in their paper implies their algorithm did not find the optimal solution. Their expected annual costs are (18.65, 17.33, 15.80, 16.33). These are different from our value (18.761, 17.428, 15.888, 16.555) which were obtained using the optimal solution to the dual of the constrained maximum entropy problem.

Using the duality based algorithm presented in this paper, we are able to go even further than Freund and Saxena and solve the original problem in Thomas. We obtain the optimal solution

$p_1^*= .25$ $p_2^* =.25$ $p_3^* =.10$ $p_4^*=.0472$ $p_5^*=.0472$ $p_6^* =.0472$
$p_7^*= .08183$ $p_8^*= .0472$ $p_9^*=.0472$ $p_{10}^*= .0274$ $p_{11}^*= .0274$ $p_{12}^*=.0274$.

This has a maximum entropy value of $H(p^*)= 2.14425549$. The solution in Thomas is different from our results even though it has only a slightly smaller $H(p^*)$ value. The resulting "optimal" probabilities differ quite a bit from our solution in some cases ($p_5^*= .044$ instead of .0472, $p_6^*= .051$ instead .0472).

In the dual formulation of the maximum entropy problem, as we proved, the computation is easily accomplished using any of a number of existing non-linear programming codes, and it is guaranteed to obtain

the optimal solution because of the special known parametric form of the primal problem, and our dual programming technique for obtaining the parameters. It can be applied in both (D) and (D'). We find that the optimal solution in (D') is very close to the optimal solution in (D) as expected, and so we might use (D') instead of (D) in certain cases if we prefer unconstrained optimization. Also we can extend the result of the Charnes-Cooper duality theory to continuous maximum entropy cases in the same manner (Charnes et al [1978]). Additionally, all of the duality results (and consequent computational savings) presented in this note on maximum entropy estimation carry over directly to minimum discrimination information (MDI) estimaiton with non-uniform "goal densities". See Charnes, Cooper and Seiford [1978] and Brockett, Charnes and Cooper [1980] for details.

REFERENCES

Brockett, P., Charnes, A., and Cooper, W.W., 1980, "MDI Estimation via Unconstrained Convex Programming," Commum. Stat. Stimula. and Computa., B9(3), pp. 709-716.

Brockett, P., Charnes, A., and Paick, K., 1984, "Computation of Minimum Cross Entropy Estimates: An Unconstrained Dual Convex Programming Method," CCS Report 491, The University of Texas at Austin.

Charnes, A., and Cooper, W.W., 1975, "Constrained Kullback-Leibler Estimation: Generalized Cobb-Douglas Balance, and Unconstrained Convex Programming," Rend. Acad. Naz. Lincei, 6, pp. 568-576.

Charnes, A., Cooper, W.W., and Seiford, L., 1978, "Extremal Principles and Optimization Dualities for Khinchin-Kullback-Leibler Estimation," Math. Operationsforsch. Statist. Ser. Optimization, 9, pp. 21-29.

Charnes, A., Cooper, W.W., and Tyssedal, J., 1983, "Khinchin-Kullback-Leibler Estimation with Inequality Constraints," Math. Operationsforsch. u. Statist., Ser. Optimization., 14, pp. 1-4.

Freud, D. and Saxena, U., 1984, "An Algorithm for a Class of Discrete Maximum Entropy Problems," Opns. Res., 32, pp. 210-215.

Guiasu, S., Martin, L., and Reischer, C., 1980, "On the Structure of Random Modular Networks," J. Comb., Inf. & Syst. Sci., 5, pp. 205-214.

Leblanc, R. and Reischer, C., 1981, "On Second Order Approximation for the Solution of the Principle of Minimum Interdependence," J. Comb., Inf. & Syst. Sci., 6, pp. 308-314.

Thomas, M.U., 1979, "A Generalized Maximum Entropy Principle," Opns. Res., 27, pp. 1188-1195.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER CCS 492	2. GOVT ACCESSION NO. AD-A159 784	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Simple Unconstrained Dual Convex Programming Method for the Computation of Discrete Maximum Entropy Distributions		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) P. Brockett, K. Paick		8. CONTRACT OR GRANT NUMBER(s) N00014-82-K-0295 N00014-81-C-0236
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Cybernetic Studies The University of Texas at Austin Austin, TX 78712		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 434) Washington, D.C.		12. REPORT DATE May 1984 13. NUMBER OF PAGES 14
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Maximum entropy computation, dual program for maximum entropy problems

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We formulate the generalized constrained maximum entropy problem often used in a decision making context as an extended dual convex programming problem. We then present the dual problem. In this dual setting the primal Lagrange multipliers are precisely the dual variables, and are easily calculated directly by virtue of the simple structure of the dual problem. An example involving the selection of best equipment for an oil spill is presented as an illustration. We contrast our solution with those given by previous authors.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

11-85

DTIC